图书情报工作
第 65 卷 第 9 期 2021 年 5 月

社交媒体虚假健康信息特征识别*

■ 张帅
武汉大学信息管理学院 武汉 430072 武汉大学信息资源研究中心 武汉 430072

摘 要:[目的/意义]识别社交媒体虚假健康信息特征,构建社交媒体虚假健康信息特征清单,以期为社交媒体虚假健康信息特征的测度提供一定理论支撑,也为用户和社交媒体平台判别虚假健康信息提供有益参考。[方法/过程]采集 1 004 条社交媒体健康数据,利用程序化编码抽取社交媒体虚假健康信息的关键特征,运用卡方检验和方差分析揭示社交媒体虚假健康信息的显著特征,并构建社交媒体虚假健康信息特征清单。[结果/结论]研究结果表明,社交媒体虚假健康信息特征具有表面特征、语义特征和来源特征 3 个维度、11 个主要特征以及 29 个子特征。其中,社交媒体上食品安全主题的虚假健康信息在"术语包装"特征上表现得更为显著;"夸大事实"为社交媒体上常见疾病主题虚假健康信息的显著特征;社交媒体上养生保健主题的虚假健康信息具有"元数据缺失"和"假借权威"显著特征。

关键词:社交媒体 健康信息 虚假特征 特征识别
分类号:G250
DOI:10.13266/j.issn.0252-3116.2021.09.008

1 引言

随着人口老龄化和疾病年轻化,寻求健康信息成为人们当下迫切的现实需求[1]。而社交媒体作为创建、发布与传播信息的重要平台,已成为人们获取健康信息的新途径[2]。根据一项针对用户搜寻健康信息方式的调查显示,63.26%的用户使用社交媒体搜索健康信息[3]。关于用户寻求健康信息渠道的偏好研究也表明,相比于其他的在线健康资讯平台,用户更加倾向于使用社交媒体来查找健康信息[4]。尽管社交媒体已成为用户获取健康信息的重要来源,但其也充斥着大量无用甚至有害的虚假健康信息[5],饱受信息质量参差不齐[6]、信息污染严重[7]、信息真假难辨[8]等诟病。最新关于虚假健康信息的调查显示,31.2%的互联网用户遭遇过虚假健康信息的欺诈[9],超过 50%的社交媒体传播的健康信息为虚假健康信息[10]。此外,皮尤研究中心(Pew Research Center)的报告也显示,68%的社交媒体用户经常看到虚假或不真实的健康信息[11]。加之社交媒体缺乏有效的监管[12]以及用户的健康信息素养水平有限[13],用户的健康决策与行为也受到虚假健康信息的误导而存在较大的风险,甚至可能导致严重的健康后果[14]。比如近期在社交媒体上流传的关于"饮用高度酒精可以预防新冠肺炎"的虚假信息严重影响了公众对新冠肺炎病毒的认知,并因此导致了至少 600 人死亡[15]。不仅如此,伴随着新型冠状病毒疫情的爆发,社交媒体上虚假健康信息的快速传播更是导致了新一轮的"信息疫情"爆发[16]。

因此,如何从社交媒体海量的健康信息中辨别出虚假健康信息是当前急需解决的现实问题[17],而识别社交媒体虚假健康信息的特征对于虚假健康信息的有效辨别至关重要[18-19]。鉴于此,笔者将聚焦于挖掘社交媒体虚假健康信息的特征,通过采集真实的社交媒体健康数据为分析样本,基于图书情报学领域的相关理论和技术,对社交媒体虚假健康信息特征识别展开深入探究,以期为社交媒体虚假健康信息特征的测度提供一定理论支撑,也为用户和社交媒体服务平台判别虚假健康信息提供有效途径。

2 概念界定及相关研究

2.1 社交媒体虚假健康信息

Web 2.0 时代,社交媒体正日益成为虚假健康信息传播的主要渠道[20]。首先,由于社交媒体的开放

* 本文系湖北文化名家专项基金资助研究成果之一。
作者简介:张帅(ORCID:0000-0002-5792-877X),博士研究生,E-mail:zs09053@163.com。
收稿日期:2020-11-15 修回日期:2021-02-02 本文起止页码:70-78 本文责任编辑:徐健

张帅. 社交媒体虚假健康信息特征识别[J]. 图书情报工作,2021,65(9):70－78.

性,任何用户都可以直接发布和传播健康信息,而无需确认其所发布信息的真实性或可靠性[21]。其次,研究也表明,用户在社交媒体上分享自身健康经验的动机往往比在其他在线平台上更为强烈,而这些第一手的健康信息可能是不准确的[22]。最后,与网站上的健康信息相比,社交媒体上的健康信息趋于简化,容易忽略一些细小但重要的信息[21]。鉴于此,笔者将重点关注社交媒体上的虚假健康信息,这里的健康信息是指一切有关人的健康的知识、技术、技能、观念和行为等信息内容[23]。根据 L. Bode 和 E. K. Vraga 的观点[24],笔者将虚假健康信息定义为缺乏科学证据和专家意见支持的错误健康信息。该定义可以用来描述社交媒体上不同类型的错误健康信息,如谣言、失真健康信息、伪健康信息及其他变体[25]。

2.2 虚假信息特征研究

随着越来越多的虚假信息在网上传播,一些学者也针对虚假信息的特征展开了相关研究。J. Zhou 等调查了 Twitter 上健康谣言的特征,发现 Twitter 上的健康谣言具有以下 7 个特征:情感效价、吸引力、发布者的权威性、外部证据、论据长短、主题标签以及直接消息[26];Y. Li 等采用 CARS 列表将网络虚假健康信息特征提炼为 4 个维度:缺乏可信度、缺乏准确性、缺乏合理性和缺乏相关支持[27];L. Rubin 将互联网上的虚假信息特征归纳为 3 个方面:伪造、隐瞒(忽略重大事实)和模凌两可[28];L. Zhou 等将网站上虚假信息特征划分为 9 类:数量、复杂性、不确定性、非直接性、表达性、多样性、非正式性、特异性以及影响力[29];L. Lavorgna 等分析了网络虚拟社区中的虚假健康新闻的特征,表明虚假健康新闻往往夸大了未经科学证实的事实[30]。虽然现有研究发现了虚假信息的一些特征,但这些特征大多是针对外文的,是否适用于中文仍有待验证,而且这些特征的提取较少来自实证数据。此外,已有研究对社交媒体上的虚假健康信息特征的关注仍然比较缺乏。

2.3 虚假信息识别研究

目前,已有一些研究试图开发评估工具以识别虚假信息。DISCERN 是由英国牛津大学健康科学研究所开发的广泛用于帮助消费者识别互联网上虚假治疗信息的一款工具,该工具包含 15 项关键指标,涉及治疗信息的明确性、相关性、适当性、参考文献、日期、客观性等[31];HONcode 是瑞士健康在线基金会建立的主要用于帮助患者、医护工作者和普通大众识别健康网站上虚假信息的一套准则,该准则包含 8 项评估原则,涉及健康信息的权威性、溯源性、合理性、透明度、广告及编辑政策等[32];LIDA 是由英国牛津大学附属公司 Minervation 开发的一款用于识别网站上虚假医疗保健信息的工具,该工具由 3 部分组成,即信息的可访问性、可用性和可靠性[33];JAMA 是由美国医学会杂志开发的一套用于识别医疗网站上虚假健康信息的标准,该标准包含 4 个维度:作者、归因、信息披露和更新,具体包括作者姓名、作者单位、授权证书、参考文献、来源、利益冲突、广告、日期等评价指标[34]。尽管已有研究开发了一些识别虚假信息的评估工具,但也可以发现,这些工具大多集中在医疗健康网站中的虚假信息辨别,是否适用于社交媒体中的虚假健康信息仍有待检验。

此外,也有一些学者基于机器学习方法来识别虚假信息。如 M. Ott 等[35]、S. Shojaee 等[36]、J. Li 等[37]运用词袋、词性和文体特征对人工编写的虚假评论和真实评论进行文本提取,构建了朴素贝叶斯(NBM)和支持向量机模型(SVM),达到 84% －89.6% 的识别准确率;N. Jindal 和 B. Liu 等[38－39]基于文体、元数据和语法等特征对亚马逊数据集采用逻辑回归模型进行文本挖掘,获得 63% －78% 的识别准确率;A. Mukherjee 等[40]在 Yelp 数据集上采用支持向量机分类器,运用文本特征获得 65.6% 和 67.8% 的识别准确率。可以看出,虽然应用神经网络、决策树和逻辑回归等方法判别虚假信息取得一定的效果,但当前研究大多仅关注了虚假信息的部分特征,而忽视了虚假信息其他方面的特征。因此,笔者将通过抓取真实的社交媒体健康数据,基于内容分析和统计分析混合方法,对社交媒体虚假健康信息的整体特征进行深入探究。

3 研究设计与实施

3.1 研究方法与工具

笔者采用非介入性的研究方法——内容分析法,借助 NVivo Plus 软件,对社交媒体虚假健康信息的特征进行提取。内容分析法是一种科学的对原始资料进行归纳和演绎的研究方法,通过对原始资料进行细粒度分析,能够比较客观、可靠地揭示资料数据的核心主题[41]。NVivo Plus 是目前主流的资料编码分析的工具,其插件 NCapture 具有采集与分析社交媒体数据的功能,特别适用于文本的数据获取与分析[42]。在提炼社交媒体虚假健康信息特征的基础上,笔者进一步结合卡方检验和方差分析方法来揭示虚假健康信息的显著特征。

3.2 研究样本选取

本研究以微信中传播的虚假健康信息作为研究样本。微信作为主流社交媒体的代表，向用户提供丰富的健康资讯和相关服务，是中国最具影响力的社交媒体之一。同时，调查发现，微信平台上超过一半的健康信息为虚假健康信息[10]。因此，以微信作为本文的研究平台具有相当的代表性和典型性。本研究借助微信辟谣助手来筛选虚假健康信息。微信辟谣助手是微信平台开发的辟谣小程序，已有超过 400 家权威机构、政府组织和主流媒体（如人民网、中科院之声、科普中国、丁香医生、果壳网）已接入该助手，其主要目的是曝光缺乏科学依据的虚假信息[43]。同时，该助手还发布经过科学检验的、被证明为真实的健康信息，以驳斥相应的虚假健康信息。因此，笔者利用 NCapture 在微信辟谣助手上采集 2018 年 1 月 –2020 年 8 月微信中被证实为虚假的健康信息，共 502 条。同期，采集微信中驳斥这些虚假健康信息的真实健康信息，共 502 条。此外，本研究排除尚不能判断其真假的健康信息。之后，笔者对共计 1 004 条真假健康信息进行文本遍历并保存为 PDF 格式，为每条真假健康信息建立资料索引，本次样本采集工作共持续两个月。

在此基础上，本研究根据世界卫生组织（World Health Organization，WHO）对健康主题的分类[44]，由两名研究者共同对每条真假健康信息的主题进行归类，最终资料主题被归为 3 类：①食品安全主题（31.3%），包括食品添加剂、转基因食品、食物搭配等；②常见疾病主题（27.7%），包括流感、疫苗、高血压、糖尿病、心脏病、艾滋病等；③养生保健主题（41.0%），包括保健调养、健身减肥、孕产护理等。由于真实健康信息与虚假健康信息是配对的，因此它们具有相同的主题，每类中真假健康信息的条数各占 50%。

3.3 资料编码

笔者基于国际图书馆协会联合会（International Federation of Library Associations and Institutions，IFLA）2016 年发布《如何分辨虚假新闻》（*How To Spot Fake News*）的信息图表[45]，构建虚假健康信息特征编码分析框架，以提取社交媒体虚假健康信息的主要特征。IFLA 开发虚假新闻识别信息图表主要是为了帮助普通大众分辨和鉴别媒体中虚假信息，从 8 个方面来考虑信息的真实性：考虑新闻来源、读“全”、查询作者信息、论据、核实日期、是不是一个玩笑、核实自己对此新闻有无偏见、请教专家。

为确保编码的准确性，笔者结合研究情境对 IFLA 虚假新闻识别信息图表中的特征进行适当归纳和总结。其中，“查询作者信息”和“核实日期”揭示的是“元数据缺失”特征，“考虑新闻来源”揭示的是“信源模糊”特征，‘读“全”’揭示的是“信息不完整”特征，“论据”揭示的是“缺乏确证”特征，“是不是一个玩笑”揭示的是“夸大事实”特征。另外，“核实自己对此新闻有无偏见”和“请教专家”不是针对虚假信息本身所提出的特征，因此本研究予以剔除。最终得出虚假信息的 5 个关键特征：信源模糊、信息不完整、元数据缺失、缺乏确证、夸大事实。因此，笔者以此开发社交媒体虚假健康信息特征编码初始框架（见表 1）。同时，由于 IFLA 虚假新闻识别信息图表是基于英语语境提出来的，在中文语境下未必完全适用。加之本文研究的是社交媒体的健康信息，与普通媒体的一般信息存在一定差异。因此，本文在编码的过程中虽以表 1 为框架，但并不局限于编码框架中已有的特征。

表 1 社交媒体虚假健康信息特征编码框架

关键特征	含义	参考来源
信源模糊	提供的信息的来源不明或来源不可靠	“考虑新闻来源”
信息不完整	提供的信息内容不全或有偏差	‘读“全”’
元数据缺失	无法提供与信息有关的各种描述性参数	“查询作者信息和核实日期”
缺乏确证	提供的信息不能支撑文中的观点	“论据”
夸大事实	提供的信息故意夸大事实	“是不是一个玩笑”

为确保研究的科学性，笔者参照 J. M. Corbin 和 A. Strauss 的程序化编码方法[46]，即按照开放式编码、主轴编码和选择式编码对虚假健康信息进行编码分析。鉴于本研究在提取虚假健康信息的特征上，需要与真实健康信息的特征进行比较分析，以确定社交媒体虚假健康信息的关键特征。因此，笔者对真实健康信息也按照相同的编码方式进行分析，且单独计入真实健康信息的特征编码。本研究采用两名编码员进行资料编码，为了确保编码的可靠性，随机选取 20% 的数据由两名编码员同时进行编码，并计算了编码员之间的一致性 Cohan’s kappa 系数为 0.82，表明编码一致性较高[47]。剩下的 80% 的数据随机分为两组分别进行编码，对编码过程中存在异议的虚假特征组织小组讨论并最终选择一个与研究主题最为贴切的虚假特征。

第一阶段（Code Ⅰ）开放式编码，研究者在 NVivo Plus 中采用逐行逐句的方式检阅每条健康信息，通过仔细审查原始资料和反复归纳，并稍加以词汇规范，提炼出 29 个初始范畴，4 478 个编码参考点，包括来源不明、内容残缺、缺少作者信息、缺乏科学依据、鼓吹功

效、语气无节制、假冒权威媒体、滥用术语、拼写错误、声称最新消息、编造案例等。第二阶段(Code Ⅱ)主轴编码,研究者围绕社交媒体虚假健康信息特征编码框架,运用子范畴与主范畴之间的逻辑关系,将 29 个初始范畴尽可能多的映射到这些特征中,最终形成 11 个主范畴。这些主范畴既包括信源模糊、信息不完整、元数据缺失、缺乏确证、夸大事实等编码框架中的特征,也包括编码框架之外的新开发的特征,如语气或语言不当、假借权威、术语包装、格式混乱、信息诱导、捏造信息。第三阶段(Code Ⅲ)选择式编码,研究者采用意义构建的方式对主范畴进行融合和聚焦[48],将 11 个主范畴归并为表面特征、语义特征和来源特征 3 个核心范畴。其中,表面特征是指信息格式的正确性,包括元数据缺失和格式混乱。语义特征是指信息内容的准确性,包括信息不完整、夸大事实、语气或语言不当、术语包装、信息诱导及捏造信息;来源特征是指信息来源的可信度,包括信源模糊、缺乏确证和假借权威。

编码可视化如图 1 所示,图 1 中不同分支颜色代表不同虚假健康信息特征的编码,分支的宽度代表编码占比(来源特征 31.2%、语义特征 49.6%、表面特征 19.2%),流向代表编码节点之间的关系,相邻两条竖线区域代表编码阶段(Code Ⅰ、Code Ⅱ、Code Ⅲ),该图能够直观地反映社交媒体虚假健康信息特征的编码过程与编码关系。

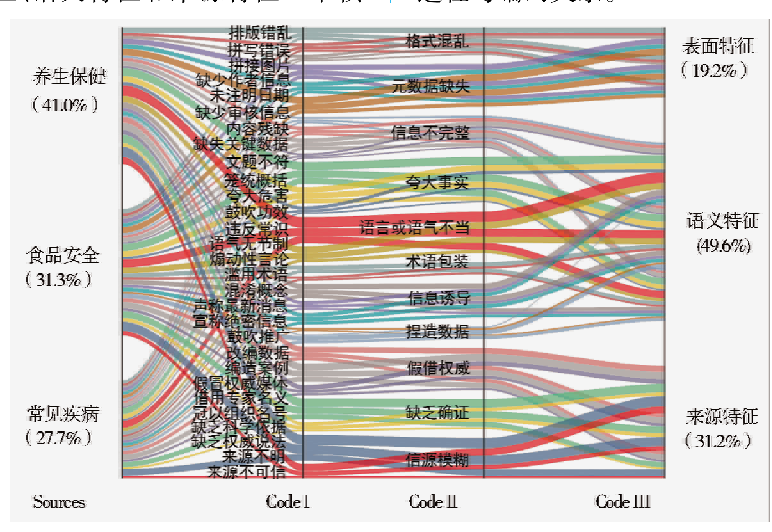


图 1　社交媒体虚假健康信息特征的编码可视化

3.4　理论饱和度检验

经过三级编码,社交媒体虚假健康信息特征已基本确定,为了确保本研究抽取的社交媒体虚假健康信息特征已经达到理论饱和,研究者按照相同的编码规则和程序连续 10 次随机选取 30 条新的虚假健康信息进行编码[49],发现没有出现新的编码范畴和编码关系,因此,可以认为本研究社交媒体虚假健康信息特征被全部捕捉到,理论饱和度较好。

4　数据分析

经过资料编码,本研究从社交媒体中提取的虚假健康信息特征具有 3 个维度、11 个主要特征、29 个子特征。为了更为精准地识别社交媒体虚假健康信息的显著特征,笔者统计了真假健康信息在不同主题和不同特征维度上的编码分布,数据汇总见表 2。表 2 中 F 代表虚假健康信息,T 代表真实健康信息,括号外数值代表虚假特征出现的频次,括号内数值代表虚假特征出现频次的占比(%)。在此基础上,本研究采用卡方检验识别相同主题中虚假健康信息的显著特征,采用方差分析比较不同主题中虚假健康信息的显著特征,具体分析如下:

4.1　相同主题虚假健康信息显著特征分析

笔者采用列联卡方检验的方法对同一主题中每个特征分别进行检验(见表 3),以测试社交媒体相同主题下真假健康信息特征之间是否具有显著差异。在本研究中,当 2 统计量所对应的近似概率值(p 值)小于显著性水平($\alpha = 0.05$),则认为社交媒体真假健康信息特征之间存在显著差异,卡方检验结果见表 3。结果表明,在食品安全类、常见疾病类和养生保健类主题中,社交媒体真假健康信息特征之间均存在显著差异,即表面特征、语义特征和来源特征及其子特征能够有效区分虚假健康信息与真实健康信息,若社交媒体健康信息在这些特征维度上均表现突出,则可以将其判定为虚假健康信息。

表 2 社交媒体真假健康信息特征编码数据汇总

特征维度	主要特征	食品安全主题($n_1=314$)		常见疾病主题($n_2=278$)		养生保健主题($n_3=412$)	
		F	T	F	T	F	T
表面特征	元数据缺失	141(89.9)	6(3.8)	129(92.8)	3(2.2)	193(93.7)	17(8.3)
	格式混乱	107(68.2)	11(7.0)	94(67.6)	6(4.3)	137(66.5)	15(7.3)
语义特征	信息不完整	104(66.2)	4(2.5)	101(72.7)	2(1.4)	138(67.0)	3(1.5)
	夸大事实	139(88.5)	5(3.2)	129(92.8)	4(2.9)	184(89.3)	7(3.4)
	语气或语言不当	145(92.4)	55(35.0)	131(94.2)	40(28.8)	190(92.2)	61(29.6)
	术语包装	50(31.8)	4(2.5)	37(26.6)	1(0.7)	47(22.8)	2(1.0)
	信息诱导	127(80.9)	6(3.8)	114(82.0)	9(6.5)	170(82.5)	15(7.3)
	捏造数据	62(39.5)	2(1.3)	52(37.4)	1(0.7)	74(35.9)	4(1.9)
来源特征	信源模糊	146(93.0)	23(14.6)	131(94.2)	18(12.9)	191(92.7)	25(12.1)
	缺乏确证	127(80.9)	7(4.5)	114(82.0)	3(2.2)	173(84.0)	7(3.4)
	假借权威	110(70.1)	3(1.9)	126(90.6)	4(2.9)	186(90.3)	6(2.9)

在确定社交媒体虚假健康信息显著特征的基础上,笔者进一步根据卡方统计量来判断虚假特征的重要程度[50]。在表面特征维度上,“元数据缺失”(229.012***–300.851***)是社交媒体虚假健康信息重要的特征,“格式混乱”(124.653***–155.167***)是社交媒体虚假健康信息的次要特征。在语义特征维度上,“夸大事实”(225.240***–305.786***)、“信息诱导”(160.763***–235.702***)、“信息不完整”(141.136***–196.506***)、“语气或语言不当”(111.553***–169.659***)是社交媒体虚假健康信息重要的特征,“捏造数据”(60.635***–77.491***)、“术语包装”(39.505***–47.324***)是社交媒体虚假健康信息次要的特征。在来源特征维度上,“信源模糊”(184.682***–268.166***)、“缺乏确证”(181.836***–271.865***)和“假借权威”(158.279***–316.023***)均是社交媒体虚假健康信息重要的特征。

表 3 社交媒体真假健康信息特征卡方检验

特征维度	主要特征	食品安全主题($n_1=314$)		常见疾病主题($n_2=278$)		养生保健主题($n_3=412$)	
		F	T	F	T	F	T
		χ^2		χ^2		χ^2	
表面特征	元数据缺失	233.111***		229.012***		300.851***	
	格式混乱	125.122***		124.653***		155.167***	
语义特征	信息不完整	141.136***		151.161***		196.506***	
	夸大事实	230.318***		225.240***		305.786***	
	术语包装	47.324***		39.505***		46.905***	
	信息诱导	190.972***		160.763***		235.702***	
	语气或语言不当	111.553***		125.819***		169.659***	
	捏造数据	70.650***		60.635***		77.491***	
来源特征	信源模糊	193.859***		184.682***		268.166***	
	缺乏确证	187.463***		181.836***		271.865***	
	假借权威	158.279***		215.060***		316.023***	

注:***表示 $p<0.001$

4.2 不同主题虚假健康信息显著特征分析

笔者采用方差分析方法对不同主题中的每个特征分别进行检验,以推断社交媒体不同主题虚假健康信息之间是否存在显著差异。在本研究中,以不同主题作为处理组(第 1 组:食品安全;第 2 组:常见疾病;第 3 组:养生保健),样本含量为每组包含虚假健康信息的条数,每个特征在每条样本中出现的频次作为观测值。当特征的均值差显著性水平 p 值小于 0.05 时,表明不同主题的虚假健康信息在该特征上存在显著差异,则认为该特征为不同主题虚假健康信息的显著特征,方差分析结果见表 4。

结果表明,不同主题虚假健康信息在大多特征上尚未表现出明显差异,表明社交媒体不同主题虚假健康信息特征较为相似。但也可以发现,食品安全(组 1)

张帅. 社交媒体虚假健康信息特征识别[J]. 图书情报工作,2021,65(9):70－78.

表 4 社交媒体虚假健康信息特征方差分析

特征维度	核心特征	ANOVA p-value (sig.)	Post hoc $P_{1,2}$	Post hoc $P_{1,3}$	Post hoc $P_{2,3}$
表面特征	元数据缺失	.000 ***	.123	.000 ***	.043 *
	格式混乱	.876	.853	.612	.747
语义特征	信息不完整	.080	.061	.829	.067
	夸大事实	.000 ***	.000 ***	.712	.039 *
	术语包装	.000 ***	.000 ***	.011 *	.186
	信息诱导	.838	.692	.561	.853
	语气或语言不当	.475	.318	.914	.269
	捏造数据	.586	.572	.303	.641
来源特征	信源模糊	.668	.496	.867	.397
	缺乏确证	.506	.693	.251	.451
	假借权威	.000 ***	.877	.000 ***	.008 **

注：*** 表示 $p<0.001$，**表示 $p<0.01$，*表示 $p<0.05$

主题在“术语包装”($p_{1,2}=.000^{***}$，$p_{1,3}=.011^{*}$)特征上更为显著。而常见疾病(组 2)和养生保健(组 3)主题在“术语包装”($p_{2,3}=.186$)特征上不存在显著差异。具体而言,社交媒体食品安全主题的虚假健康信息在“术语包装”方面,经常使用“激素催生技术”“偶氮二甲酰胺”“阿斯巴甜”等专业术语或伪概念,造成发布的内容具有专业性的假象来迷惑用户。常见疾病(组 2)主题在“夸大事实”($p_{1,2}=.000^{***}$，$p_{2,3}=.039^{*}$)特征上更加显著。而食品安全(组 1)和养生保健(组 3)主题在“夸大事实”($p_{1,3}=.712$)特征上不存在显著差异。具体而言,社交媒体常见疾病主题的虚假健康信息在“夸大事实”方面,多见的表述有“震撼全球”“第一奇方”“抗癌之星”“根治”等用语宣称药品具有极高治愈率。养生保健(组 3)主题在“元数据缺失”($p_{1,3}=.000^{***}$，$p_{2,3}=.043^{*}$)和“假借权威”($p_{1,3}=.000^{***}$，$p_{2,3}=.008^{**}$)特征上更加显著。而食品安全(组 1)和常见疾病(组 2)主题在“元数据缺失”($p_{1,2}=.123$)和“假借权威”($p_{1,2}=.877$)特征上不存在显著差异。具体而言,社交媒体养生保健主题的虚假健康信息在“元数据缺失”方面,缺少“作者姓名”“发布日期”“审核者”“版权声明”等重要信息;在“假借权威”方面,经常使用“诺贝尔奖得主”“首席科学家”“院士”等名义进行保健品的虚假宣传,以欺骗消费者。

4.3 社交媒体虚假健康信息特征清单构建

通过卡方检验和方差分析,本研究揭示了社交媒体虚假健康信息的显著特征。基于上述显著特征,笔者进一步构建了社交媒体虚假健康信息特征清单,以期为用户更好地识别虚假健康信息提供有益参考,如表 5 所示。该清单主要由社交媒体虚假健康信息的特征维度、主要特征、子特征以及相关建议组成。应用上述关键特征及相关建议,用户可以对社交媒体健康信息进行综合研判。

5 研究结论与启示

笔者采集了来自社交媒体网络中的 1 004 条实证数据,利用内容分析方法提取了虚假健康信息的 11 个主要特征,并将其融合和聚类为表面特征、语义特征和来源特征 3 个维度。其中,“元数据缺失”“夸大事实”“信息诱导”“信息不完整”“语气或语言不当”“信源模糊”“缺乏确证”“假借权威”是社交媒体虚假健康信息重要的特征,“格式混乱”“捏造数据”“术语包装”是社交媒体虚假健康信息次要的特征。在此基础上,笔者运用卡方检验识别出表面特征、语义特征和来源特征为相同主题中社交媒体虚假健康信息的显著特征,采用方差分析推断出“元数据缺失”“夸大事实”“术语包装”和“假借权威”为不同主题社交媒体虚假健康信息的显著特征。其中,社交媒体上食品安全主题的虚假健康信息在“术语包装”特征上表现得更为显著;“夸大事实”为社交媒体上常见疾病主题虚假健康信息的显著特征;社交媒体上养生保健主题的虚假健康信息具有“元数据缺失”和“假借权威”显著特征。这些特征的揭示将有助于不同主题的虚假健康信息的自动识别。最终,笔者构建了社交媒体虚假健康信息特征清单,该清单不仅可以为社交媒体虚假健康信息特征的测度提供一定理论支撑,也可以为社交媒体构建虚假健康信息自动过滤的机制。

与已有研究相比[6,51－52],笔者在理论方面进行了以下 3 方面的拓展:①以往研究大多采用介入性研究方法,如访谈法、专家咨询法和问卷调查法来评估网络健康信息的特征,这些方法在不同程度上介入了研究对象的活动,难以客观反映真实的评估结果,而笔者采用非介入性研究方法,对社交媒体健康信息的真实数据进行科学地分析,形成了比较可靠的社交媒体虚假健康信息特征清单;②本文聚焦于社交媒体虚假健康信息的语法、语义和语用等特征,为社交媒体虚假健康信息特征的测度提供了情报学的研究视角和思路;③本文探索了社交媒体环境下不同主题虚假健康信息的特征,发现“元数据缺失”“夸大事实”“术语包装”和“假借权威”等为不同主题虚假健康信息的显著特征,这在以往研究中少有关注。

在实践方面,本研究不仅为社交媒体用户辨别虚

表 5　社交媒体虚假健康信息特征清单

特征维度	主要特征	子特征	相关建议
表面特征	元数据缺失	缺少作者信息 未注明日期 缺少审核信息	检查是谁发布的,什么时间发布的,审核者是谁,有无版权声明等
	格式混乱	拼接图片 排版错乱 拼写错误	注意可疑的图片,检查文章排版,检查文字拼写
语义特征	信息不完整	内容残缺 缺失关键数据 文题不符 笼统概括	小心以偏概全,断章取义,不要只读标题,要浏览全文
	夸大事实	鼓吹功效 违反常识 夸大危害	注意夸张描述词汇,如“震撼全球”“第一奇方”“抗癌之星”“根治”
	术语包装	滥用术语 混淆概念	检查专业术语,注意偷换概念
	信息诱导	声称最新消息 宣称绝密信息 诱导推广	提防诱导性的文字,如“最新发现”“紧急播报”“请立即传出去”“快告诉家里人”
	语气或语言不当	语气无节制 煽动性言论	注意文章的语言风格和语气强弱
	捏造数据	改编数据 编造案例	仔细确认数据,提防虚假案例
来源特征	信源模糊	来源不明 来源不可信	检查链接和来源,追本溯源
	缺乏确证	缺乏科学依据 缺乏权威说法	提防主观臆断,小心求证
	假借权威	假冒权威媒体 借用专家名义 冠以组织名号	搜索其他媒体是否报道过,提防谣言变种

假健康信息提供了有效工具,也为社交媒体服务提供商判别虚假健康信息提供了有效途径。对社交媒体用户而言,一方面,本文虚假健康信息特征的命名源自于相关原始语句,较为直观地揭示了虚假健康信息的特征,社交媒体用户可以借鉴虚假健康信息特征清单,对搜寻或偶遇到的健康信息进行主动识别,以减少因误信虚假健康信息而导致严重医疗后果的情况发生;另一方面,用户可以通过对虚假健康信息特征的学习和掌握,以提升自身健康信息素养,从而提高虚假健康信息鉴别能力。对社交媒体服务服务提供商而言,笔者构建的社交媒体虚假健康信息特征清单,既为社交平台管理者筛选和剔除虚假健康信息提供实证支持,也为社交媒体平台基于该框架建立虚假健康信息自动预警和过滤机制提供理论支撑,从而实现从源头上遏制虚假健康信息的传播。同时,社交媒体相关健康信息的发布者也可以参考本列表来完善其所传播的健康信息的表面、语义和来源特征,从而全面提升社交媒体平台健康信息的质量。

尽管本研究在理论和实践方面做出了一定贡献,但仍然存在一些局限。首先,笔者构建的社交媒体虚假健康信息特征清单还处于初步探索阶段,鉴于社交媒体健康信息的动态性,该列表在未来实际应用中也应不断改进和优化。其次,本研究是基于微信平台进行的探索,结论是否使用于其他社交媒体平台,还有待进一步的比较研究。最后,本研究仅关注社交媒体虚假健康信息的特征识别,未考虑其识别效果,在未来的研究中,将考虑基于这些虚假信息特征对社交媒体健康信息数据集进行机器学习,以更为精准地识别社交媒体虚假健康信息,进一步深化和推广本文的研究结论。

致谢:感谢马费成教授对本文选题、撰写及最终定稿的指导和帮助!

参考文献:

[1] 金晓玲, 冯慧慧, 周中允. 微信朋友圈中健康信息传播行为研究[J]. 管理科学, 2017, 30(1): 73 –82.

[2] 徐孝婷, 赵宇翔, 朱庆华. 在线健康社区老年用户健康信息需求实证研究[J]. 图书情报工作, 2019, 63(10): 87 –96.

[3] ZHANG X, WEN D, LIANG J, et al. How the public uses social media WeChat to obtain health information in china: a survey study [J]. BMC medical informatics and decision making, 2017, 17 (2):71 –79.

[4] CHOU W Y S, HUNT Y M, BECKJORD E B, et al. Social media use in the united states: implications for health communication [J]. Journal of medical internet research, 2009, 11(4): e48.
[5] 侯筱蓉, 付扬, 陈娟. 基于微信平台的健康信息用户感知和效用研究[J]. 现代情报, 2016, 36(10): 89 - 93.
[6] 李月琳, 张秀, 王姗姗. 社交媒体健康信息质量研究:基于真伪健康信息特征的分析[J]. 情报学报, 2018(3): 294 - 304.
[7] 莫祖英, 马费成. 网络环境下信息资源质量控制的博弈分析[J]. 情报理论与实践, 2012, 35(8): 26 - 30.
[8] GOLDACRE B. Media misinformation and health behaviors[J]. The lancet oncology, 2009, 10(9): 848.
[9] 中国互联网络信息中心. 第 44 次中国互联网络发展状况统计报告[EB/OL]. [2020 - 08 - 21]. http://www.cnnic.net.cn/hlwfzyj/hlwxzbg/hlwtjbg/201908/P020190830356787490958.pdf.
[10] 李月琳, 张秀. 大学生社交媒体健康信息甄别能力研究[J]. 图书情报知识, 2018(1): 66 - 77.
[11] PEW RESEARCH CENTER. Mobile technology and its social impact survey 2018[EB/OL]. [2020 - 08 - 21]. https://www.pewresearch.org.
[12] ZHANG Z, ZHANG Z, LI H. Predictors of the authenticity of internet health rumors[J]. Health information & libraries journal, 2015, 32(3): 195 - 205.
[13] DIVIANI N, PUTTE B, GIANI S, et al. Low health literacy and evaluation of online health information: a systematic review of the literature[J]. Journal of medical internet research, 2015, 17(5): e112.
[14] 宋士杰, 赵宇翔, 宋小康, 等. 互联网环境下失真健康信息可信度判断的影响因素研究[J]. 中国图书馆学报, 2019, 45(4):72 - 85.
[15] 袁建霞, 冷伏海. "信息疫情"应对中的情报调研和分析[J]. 竞争情报, 2020, 16(3): 9 - 13.
[16] 王世伟. 略论"信息疫情"十大特征[J]. 图书馆杂志, 2020, 39(3):19 - 23.
[17] DEVINE T, BRODERICK J, HARRIS L M, et al. Making quality health websites a national public health priority: toward quality standards[J]. Journal of medical internet research, 2016, 18(8): e211.
[18] 邓胜利, 汪奋奋. 互联网治理视角下网络虚假评论信息识别的研究进展[J]. 信息资源管理学报, 2019(3): 73 - 81.
[19] 吴佳芬, 马费成. 产品虚假评论文本识别方法研究述评[J]. 数据分析与知识发现, 2019, 3(9): 1 - 15.
[20] PAL A, CHUA A Y K, GOH D H L. Does KFC sell rat? analysis of tweets in the wake of a rumor outbreak[J]. Aslib journal of information management, 2017, 69(6): 660 - 673.
[21] BRADY J T, KELLY M E, STEIN S L. The trump effect: with no peer review, how do we know what to really believe on social media? [J]. Clinics in colon and rectal surgery, 2017, 30(4): 270 - 276.
[22] FICHMAN R G, KOHLI R, KRISHNAN R. The role of information systems in healthcare: current research and future trends[J]. Information systems research, 2011, 22(3): 419 - 428.
[23] TSAI H, BAGOZZI R P. Contribution behavior in virtual communities: cognitive, emotional, and social influences[J]. Management information systems quarterly, 2014, 38(1): 143 - 164.
[24] BODE L, VRAGA E K. In related news, that was wrong: the correction of misinformation through related stories functionality in social media[J]. Journal of communication, 2015, 65(4): 619 - 638.
[25] TANDOC JR E C, LIM Z W, LING R. Defining "fake news" a typology of scholarly definitions[J]. Digital journalism, 2018, 6(2): 137 - 153.
[26] ZHOU J, LIU F, ZHOU H. Understanding health food messages on Twitter for health literacy promotion[J]. Perspectives in public health, 2018, 138(3): 173 - 179.
[27] LI Y, ZHANG X, WANG S. Fake vs. real health information in social media in China[J]. Proceedings of the Association for Information Science and Technology, 2017, 54(1): 742 - 743.
[28] RUBIN V L. On deception and deception detection: content analysis of computer-mediated stated beliefs[J]. Proceedings of the Association for Information Science and Technology, 2011, 47(1): 1 - 10.
[29] ZHOU L, BURGOON J K, NUNAMAKER J F, et al. Automating linguistics-based cues for detecting deception in text-based asynchronous computer-mediated communications[J]. Group decision and negotiation, 2004, 13(1): 81 - 106.
[30] LAVORGNA L, DE STEFANO M, SPARACO M, et al. Fake news, influencers and health-related professional participation on the Web: a pilot study on a social-network of people with Multiple Sclerosis[J]. Multiple sclerosis and related disorders, 2018, 25(10): 175 - 178.
[31] DUEÑAS-GARCIA O F, KANDADAI P, FLYNN M K, et al. Patient-focused websites related to stress urinary incontinence and pelvic organ prolapse: a DISCERN quality analysis[J]. International urogynecology journal, 2015, 26(6): 875 - 880.
[32] BOYER C, DOLAMIC L. Automated detection of HONcode website conformity compared to manual detection: an evaluation[J]. Journal of medical internet research, 2015, 17(6): e135.
[33] TAVARE A N, ALSAFI A, HAMADY M S. Analysis of the quality of information obtained about uterine artery embolization from the internet[J]. Cardiovascular and interventional radiology, 2012, 35(6): 1355 - 1362.
[34] WINKER M A, FLANAGIN A, CHI-LUM B, et al. Guidelines for medical and health information sites on the internet: principles governing AMA Web sites[J]. Jama, 2000, 283(12): 1600 - 1606.
[35] OTT M, CHOI Y, CARDIE C, et al. Finding deceptive opinion spam by any stretch of the imagination[C] //Proceedings of the 49th annual meeting of the NAACL-HLT. Portland: association for computational linguistics, 2011: 309 - 319.
[36] SHOJAEE S, MURAD M A A, BIN AZMAN A, et al. Detecting

deceptive reviews using lexical and syntactic features[C]//International conference on intelligent systems design and applications. selangor, Malaysia: IEEE, 2013: 53－58.
[37] LI J, OTT M, CARDIE C, et al. Towards a general rule for identifying deceptive opinion spam[C]//Meeting of the Association for Computational Linguistics. Baltimore: Association for Computational Linguistics, 2014: 1566－1576.
[38] JINDAL N, LIU B, LIM E P. Finding unusual review patterns using unexpected rules[C]//Proceedings of the 19th ACM international conference on information and knowledge management. Toronto: ACM, 2010: 1549－1552.
[39] JINDAL N, LIU B. Opinion spam and analysis[C]//Proceedings of the 2008 international conference on web search and data mining. Stanford: ACM, 2008: 219－230.
[40] MUKHERJEE A, VENKATARAMAN V, LIU B, et al. What yelp fake review filter might bedoing? [C]//Proceedings of the seventh international AAAI conference on weblogs and social media. ellevue, Washington: AAAI Press, 2013: 1－10.
[41] HSIEH H F, SHANNON S E. Three approaches to qualitative content analysis[J]. Qualitative health research, 2005, 15(9): 1277－1288.
[42] QSR INTERNATIONAL PTY LTD. What is nvivo[EB/OL]. [2020－08－26]. https://www.qsrinternational.com/nvivo/what-is-nvivo.
[43] 刘本莹. 基于传染病模型的微信谣言传播及免疫动力学研究[D]. 温州: 温州大学, 2019.
[44] WORLD HEALTH ORGANIZATION. Health-topics[EB/OL]. [2020－08－26]. https://www.who.int/health-topics/.
[45] INTERNATIONAL FEDERATION OF LIBRARY ASSOCIATIONS AND INSTITUTIONS. How to spot fake news[EB/OL]. [2020－08－26]. https://www.ifla.org/node/11175.
[46] CORBIN J M, STRAUSS A. Grounded theory research: procedures, canons, and evaluative criteria[J]. Qualitative sociology, 1990, 13(1): 3－21.
[47] LANDIS J R, KOCH G G. The measurement of observer agreement for categorical data[J]. Biometrics, 1977, 33(1): 159－174.
[48] 弗利克. 质性研究导引[M]. 孙进, 译. 重庆:重庆大学出版社, 2011: 248－256.
[49] FRANCIS J J, JOHNSTON M, ROBERTSON C, et al. What is an adequate sample size? operationalising data saturation for theory-based interview studies[J]. Psychology and health, 2010, 25(10): 1229－1245.
[50] ZHA X, YANG H, YAN Y, et al. Exploring the effect of social media information quality, source credibility and reputation on informational fit-to-task: moderating role of focused immersion[J]. Computers in human behavior, 2018, 79(2): 227－237.
[51] 吕亚兰, 侯筱蓉, 黄成, 等. 泛在网络环境下公众网络健康信息可信度评价指标体系研究[J]. 情报杂志, 2016, 35(1): 196－200.
[52] EYSENBACH G, KÖHLER C. How do consumers search for and appraise health information on the world wide web? qualitative study using focus groups, usability tests, and in-depth interviews[J]. British medical journal, 2002, 324(7337): 573－577.

Study on Feature Identification of False Health Information on Social Media

Zhang Shuai
School of Information Management, Wuhan University, Wuhan 430072
Center for Studies of Information Resources, Wuhan University, Wuhan 430072

Abstract: [Purpose/significance] This study identifies the features of false health information on social media and construct a list of false health information characteristics on social media, in order to provide certain theoretical support for the measurement of false health information features on social media, and also provide a useful reference for users and social media service platforms to distinguish false health information. [Method/process] 1 004 pieces of empirical data from social media were collected, and the key features of false health information were extracted by programmatic coding. Then the chi - square test and analysis of variance were adopted to identify significant features of health misinformation. In addition, this study developed a list of features to identify health misinformation on social media. [Result/conclusion] It was shown that the features of false health information on social media had three dimensions: surface features, semantic features, and source features. There were 11 main features and 29 sub-features. It was found that the features of "term packaging" on food safety topic was more notable than other topics; "exaggerated facts" on common diseases topic was more significant than other topics; the features of "lack of meta-information" and "fake authority" on healthcare topic were more prominent than other topics.

Keywords: social media health information false features feature identification